



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Back-ups for the future

**Citation for published version:**

Currie, ME & Paris, BS 2018, 'Back-ups for the future: Archival practices for data activism', *Archives and Manuscripts*, vol. 46, no. 2, pp. 124-142. <https://doi.org/10.1080/01576895.2018.1468273>

**Digital Object Identifier (DOI):**

[10.1080/01576895.2018.1468273](https://doi.org/10.1080/01576895.2018.1468273)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Archives and Manuscripts

**Publisher Rights Statement:**

This is an Accepted Manuscript of an article published by Taylor & Francis in [JOURNAL TITLE] on 12 Jun 2018, available online: <https://www.tandfonline.com/doi/full/10.1080/01576895.2018.1468273>.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Back-ups for the Future: Archival Practices for Data Activism

## Acknowledgements

Special thanks to Joan Donovan, volunteers for the Environmental Data Governance Initiative, and Stanford's Digital Civil Society Lab.

## Authors

Dr. Morgan E. Currie<sup>i</sup>  
Digital Civil Society Lab  
Center on Philanthropy & Society  
Stanford University  
Stanford Law School  
559 Nathan Abbott Way  
Stanford, CA 94305  
605 724 2654

Dr. Morgan E. Currie is a Lecturer in Data & Society, Science, Technology & Innovation Studies at the University of Edinburgh. Currie's research examines the way cultural, political, and economic factors interact with the design and development of data infrastructures, with a focus on government and civic data. Currie earned a Ph.D. in Information Studies from University of California, Los Angeles.

Britt S. Paris  
Doctoral Candidate  
Department of Information Studies  
University of California, Los Angeles  
GSE&IS Building, Box 951520  
Los Angeles, California 90095  
United States of America  
310 825 8799  
parisb@ucla.edu

Britt S. Paris is a Doctoral Candidate in the Department of Information Studies at the University of California, Los Angeles. Paris works in the field of critical informatics; she engages in fieldwork and critical discourse analysis to study how groups understand, build, and use Internet infrastructure according to their value systems.

## Abstract

Literature on activist archiving theorizes the power of recordkeeping to give voice to marginalised communities. However, missing from this archival literature are analyses about the political practice of preserving *data* as an act of grassroots resistance. Simultaneously, existing scholarly literature on grassroots data activism revolves around the creation of new statistical representations to challenge official ones. This literature has largely ignored what will happen to this data over the long-term, nor has it treated data archiving as an activist project in its own right. This theoretical article seeks to close the gap between literature on archival activism and literature on data activism, in hopes that both sets of research can draw productively from each other. There are clear affinities between activist archives and

data activism: both address the failure by mainstream institutions to account for marginal voices, both have the power to make issues visible and legitimate within the public sphere, and both experiment with traditional forms of memory and statistical evidence. We believe that these two powerful forms of activity have much to learn from each other, particularly as the need to steward data over the long term will only grow.

**Keywords:** Data activism, archival activism, community archives, critical data studies

**Disclosure statement**

No potential conflict of interest was reported by the authors

---

<sup>i</sup> Morgan Currie is currently Lecturer in Data & Society, Science, Technology & Innovation Studies, at the University of Edinburgh. Address: STIS, University of Edinburgh, 2.05 Old Surgeons' Hall, High School Yards, Edinburgh, United Kingdom, EH1 1LZ. Tel: 44 (0) 131 6517112. Email: morgan.currie@ed.ac.uk.

# Back-ups for the Future: Archival Practices for Data Activism

## Abstract

Literature on activist archiving theorizes the power of recordkeeping to give voice to marginalised communities. However, missing from this archival literature are analyses about the political practice of preserving *data* as an act of grassroots resistance. Simultaneously, existing scholarly literature on grassroots data activism revolves around the creation of new statistical representations to challenge official ones. This literature has largely ignored what will happen to this data over the long-term, nor has it treated data archiving as an activist project in its own right. This theoretical article seeks to close the gap between literature on archival activism and literature on data activism, in hopes that both sets of research can draw productively from each other. There are clear affinities between activist archives and data activism: both address the failure by mainstream institutions to account for marginal voices, both have the power to make issues visible and legitimate within the public sphere, and both experiment with traditional forms of memory and statistical evidence. We believe that these two powerful forms of activity have much to learn from each other, particularly as the need to steward data over the long term will only grow.

Keywords: Data activism, archival activism, community archives, critical data studies

## Introduction: Archival Activism Meets Data Activism

After Donald Trump, an avowed climate change denier, was elected president of the United States, a groundswell of civic activity surfaced in response to fears that federal data related to climate change would be removed from United States government webpages. Groups of students, researchers, activists, and librarians came together in cities across the country to participate in ‘DataRescues’ organised by the Environmental Data Governance Initiative (EDGI), an international civil society organization focused on threats to federal environmental and energy policy. At DataRescues participants learned how to ‘harvest,’ ‘scrape,’ and archive federal websites and data sets related to climate change, all in an effort to address potential dangers to the climate protection goals set by the global community in the last 40 years and to the mainstream science that investigates how humans are changing the planet.

DataRescues have created an innovative archival model by leveraging distributed networks of volunteers to crowdsource data archiving work. DataRescues focus on the long-term

custody of a numerical dataset or multiple datasets; they explicitly deploy archival principles to safeguard data, and their participants have designed software with archival principles in mind, such as chain of custody, provenance, checksums, and multiple copies. These events are examples of what we have elsewhere called ‘activist data archiving;’ they entail a novel type of activism in the form of concerned citizens who come together to archive data of political import.<sup>1</sup>

To better understand *data archiving* as a particular form of political activism, we draw from two sub-disciplines—archival activism and data activism—that to-date have been separately theorised in academic literature. Literature on activist and community archives, for its part, analyses archives as political projects that are often integral in the struggle for human rights and social justice.<sup>2</sup> Community archives, more specifically, offer marginalised people and social movement actors, such as racial and ethnic minorities and LGBTQ communities, the chance to tell their own accounts of the past.<sup>3</sup> By creating the physical infrastructure needed to steward evidence and memories over time, activist archives preserve the political and cultural power of these collected records.

Data archiving can also be situated within a growing body of literature on data activism. At present, we commonly associate ‘data’ with the technology industry’s vast digital collection of personal, biological, climatic, geographic, transactional and economic phenomena or events. Research on data activism looks instead at grassroots groups that collect, process, and visualise data themselves, often in statistical form. Because society typically perceives numerical or statistical data as unbiased and authoritative, statistical accounts offer civic groups a persuasive form of evidence to present to governments and to the public. Simultaneously, data activism foregrounds the powerful political dimensions of data as a form of representation. Data activism

– also called civic data, or statactivism in the academic literature – makes clear the political, economic, and cultural power of data, particularly as data flows are now such a part of daily life.

There are clear affinities between archival activism and data activism. Both forms focus on including interested stakeholders in acts of political representation that offer unique perspectives on political or social justice. The disparate literatures on data and archival activists both draw from disciplines – including critical data studies, archival studies, and information studies – that reveal the culturally-laden practices underlying information production and use.<sup>4</sup> Even the differences between records and data themselves are sometimes not clear cut.<sup>5</sup> A dataset can act as a set of records, and records can take the form of statistical data; both can play an evidentiary function.

However, the literature on data activism has widely ignored issues of long-term data stewardship and the question of how largely ad hoc, voluntary community groups can maintain the datasets' long-term integrity and sustainability. Scholarship on data activism tends to focus on civic data *collection* to create new statistical representations or on data *analysis* that challenges official statistical accounts. Less scholarship has focused on the entire data lifecycle or data *archiving* as a political act in and of itself. Archival scholars, on the other hand, theorise records with these principles in mind. This paper therefore seeks to reconcile the differences in these related literatures and highlight important areas where these disciplines can benefit one another. Thinking through both of these practices can also help us refine our understanding of data archiving as a political act.

This theoretical article sets out to close the gap between literature on activist archives and literature on data activism, in hopes that both bodies of research can draw productively from each other. We begin by introducing the distinct literatures on archival theory and critical data

studies, then make several insights into characteristics shared between archival and data activism:

1. both highlight that neither archiving nor data are politically neutral pursuits,
2. both address issues of representation and power,
3. although their participants may include non-experts, both types of activism can still be trustworthy and make legitimate claims,
4. participants of both types of activism engage in community formation,
5. both persist independently of traditional institutional contexts.

To illustrate these points further, we focus on a case study of EDGI's archival protocols and workflows developed by their members. In our conclusion, we illuminate some of the overlapping interests of scholars, activists, and community members involved in both data activism and activist archives, and we draw some lessons for data activists from archival discourses. We argue that the work of data archiving as an activist project can create new models for long-term digital stewardship, particularly by leveraging distributed networks of people who gather around a shared issue of concern.

### **Archiving for Radicals**

Within archival theory, the term 'activist archiving' describes archives that are deliberately political, progressive, and maintained by and for the very people who generate them.<sup>6</sup> As Flinn and Alexander explain, this work can happen in more ways than one; what they term *archival activism* can describe an archive - which may or may not be overseen by professionals - that collects the work of social movements, minority communities, or human or civil rights activists and often promotes stakeholder participation. *Activist archives* are repositories created by activists themselves who engage in archival description, collection, curation as a core part of

their social movement work.<sup>7</sup> Both types of practice, which often overlap, inform archival description and representation, collection and curation, with the goal of extending political power of under-represented and oppressed identities.

Grassroots archives often exist independently of large educational, governmental or private institutions; they persist, rather, based on the desire to create a collective identity and to challenge the dominant narratives of the past. Such is often the case with community archives, which offer certain groups – such as the LBGTQ population, religious or ethnic minorities, indigenous peoples, and labor movements – control over their historical representations.<sup>8</sup> Community archives offer ‘local, autonomous spaces for alternative historical narratives and cultural identities to be created and preserved;’ they often form in response to cultural silencing and marginalization, and their participants have a palpable connectedness to a shared history and cultural legacy.<sup>9</sup> The Mazer Lesbian Archive, for example, began in a private residence in Los Angeles in the 1980s and was dedicated to documenting lesbian culture, which was largely invisible at the time.<sup>10</sup> Likewise the Black Cultural Archives, a non-profit in the U.K, originally formed as a community project to collect and document the cultural history of African and Afro-Caribbean people – stories that had been largely unrecognized within the UK’s school curriculums.<sup>11</sup> Participation by a community to collect and appraise its own history is crucial to these endeavors.

While community and activist archiving provide templates for data archiving as a political act, the literature on these practices does not comprehensively address technical and political concerns that arise from archiving data in networked environments, where activist data projects often exist. Although an adjacent set of literature on data preservation *does* examine the affordances of digital data, this body of research is largely focused on preserving databases in the



institutional contexts of academic science and government. This literature covers many of the technological concerns that activist archiving research leaves out, but it does not address particular, political dimensions of data, nor offer accounts of data archiving by activist groups.<sup>12</sup> For these theoretical concerns, we can turn to literature on critical data studies and data activism.

### **Data Activism**

In recent years, scholars across fields ranging from geography to information and media studies have analysed data practices as a form of activism and resistance. In this literature, ‘data’ describes units of information that create powerful representations of phenomena in the world.<sup>13</sup> Civic data collection and analysis, variously labelled as counter-data action,<sup>14</sup> statactivism,<sup>15</sup> or data activism,<sup>16</sup> includes citizens who use devices to detect signs of radioactivity in Fukushima, Japan,<sup>17</sup> grassroots efforts to capture more comprehensive statistics on those killed and injured by police,<sup>18</sup> and citizens who use Bellingcat.com, an open-sourced ‘social media weapon tracking’ used for the Syrian conflict and which gained the praise of Amnesty International for calling attention to data privacy concerns.<sup>19</sup>

The need for data activism is increasingly clear in the contemporary world. As we tap into corporate-owned data flows for crucial life activities or generate data with each click on our multiple, connected devices, we contribute to corporate and government surveillance systems and technology industry profit. The rise of political activity surrounding datafication has come about, at least in some degree, as a response to the 2014 Snowden revelations, proving that with help from Silicon Valley, the government spies on citizens both inside the U.S. and across the globe.<sup>20</sup> With instruments such as GIS, affordable sensors, and online survey tools, communities have begun conducting their own research to influence scientific debates and public policy and to offer alternative narratives than those promoted by government or industry.<sup>21</sup>

Critical geographers Dalton and Thatcher call such acts of resistance to politically dominant datasets ‘counter data action.’<sup>22</sup> This notion draws from their work in critical GIS (geographic information systems), which asserts that maps are political and legal claims on reality, rather than uncomplicated, objective cartographic tools per conventional understanding. Based on this framework, purveyors of Public Participatory GIS engage in ‘counter mapping’ as a method of emancipatory action performed by communities, such as indigenous peoples looking to reclaim resources from external dominance. Just as counter-mapping repurposes tools of state-making to express the perspectives of the indigenous, counter-data actions combat powerful interests who insist that technology is simply neutral. Dalton and Thatcher argue that counter-data actions can offer possibilities for political liberation or aesthetic expression that actively confront or redress the power asymmetries found in technologies of surveillance or capitalist accumulation.

In a similar vein, ‘statactivism’ is a term Bruno et. al. have used to describe ‘forms of emerging collective action that use numbers, measurements and indicators as means of denunciation and criticism.’<sup>23</sup> Statactivism deploys acts of appropriation and intervention by either decrying certain authoritative metrics or devising new ones. Groups might collect and deploy their own data to make a cause more visible, such as AIDS activists in the 1980s — particularly those involved in Project Inform — who conducted their own clinical trials and epidemiological risk assessment studies.<sup>24</sup> By presenting data collected through alternative methods or from different sources, all which presume to shed light on the same phenomena, the contingent and negotiated aspects of mainstream data come to the fore. Local statistical practices in this way offer another aggregate reality that diverges from or augments the official version.

Regardless of the specific mode of data intervention, some have critiqued alternative data practices as another tool of power and rationalization appropriated from the bureaucratic state. Critics have denounced data and statistics wholesale as oppressive tools of positivist accounting that perpetuate classism, racism, and sexism.<sup>25</sup> Scholars have also called algorithms ‘inherently fascistic’ because they masquerade as a neutral alternative to human decision-making, a black box that settles human affairs through automation.<sup>26</sup> In these critiques, data-driven solutions threaten to put the onus on the most vulnerable citizens to prove and address unfair treatment while removing culpability from powerful entities responsible for these injustices.<sup>27</sup> Yet to criticise, on principle, the rationalizing logic of accounting and statistical assessment would continue to ‘allow a monopoly of these instruments to the powerful’ and sweeps aside a long history of using metrics for progressive goals.<sup>28</sup>

This existing scholarly literature on data activism, in any case, primarily revolves around the *creation* of new statistical representations or metrics that challenge official ones. The literature cited has largely ignored concerns for data over the long-term, and has not treated data archiving as a specific type of activist data project. We contend that archivists can offer critical lessons to data activists, and that they can themselves play an important role in the activist sphere in the increasingly data-driven world.

### **Data Archiving as Activism**

In the following sections we discuss overlapping concerns, methods, and orientations common to both data activism and archival activism. Analyzing these features can illuminate each disparate stream of research and help us understand new forms of activism that focus on data archiving as a political act:

#### ***1. Both Data and Archives Are Political***

In the past decades, archival theory has taken a post-positivist turn by asserting the political dimensions of archives. In 1977, Howard Zinn termed archiving ‘an inevitably political craft,’ underscoring that archiving is never neutral, as it always involves acts of interpretations, which are themselves laden with the cultural and political values of the archivist and archiving institution.<sup>29</sup> Zinn’s statement confronts a long history of archival theory that defined the record as an impartial, passive bi-product of administrative activity; in this tradition, archival activity entailed preserving the record in its original context as faithful, objective evidence of this past. By preserving this evidence, the archivist him or herself became a ‘selfless devotee of Truth’ and the steward of official history.<sup>30</sup>

The postmodern or post-positivist turn, in contrast, identifies the social influences that produce a record and underlie its archival power. Any act of preservation entails the ‘privileging of certain records and records creators, certain functions, activities, and groups in society, and the marginalizing or silencing of others,’ writes archivist Terry Cook.<sup>31</sup> Archival scholars such as Cook and Ciaran Trace emphasise the role of archives to produce cultural memory, and the political function that archival selection plays in generating ‘official’ history.<sup>32</sup> In addition to selection, the classifications and descriptions authoritatively imposed on records by past archivists play a role in how we understand records in the present day. Electronic and web archives raise further, socially urgent questions about selection and memory-making amidst the flood of digital content, along with very fraught, political and social questions about ownership of our cultural heritage on for-profit, corporately-owned platforms or servers.

Archival activism is not simply a way of re-politicising archives - archives have always already been political. What archival activism does is explicitly confront the dominant narratives of official, institutional archives and the elevation of certain people, institutions, and professions

to shape our collective memory. Archival activism goes hand-in-hand with the post-positive turn in archives by pluralizing the number of voices who shape historical narratives and civil discourse.

Critical data studies mirrors this postmodern turn in archival studies. Parallel to asserting the political nature of records keeping, data scholars in the STS tradition emphasise that data is never a neutral, passive presentation of the world but the result of a contingent set of decisions and institutional arrangements that provide a basis for their value as agreed-upon objects. Rather than question whether data reflect reality with objectivity, these thinkers claim that reality is the product of established conventions that have been widely invested in through consistent measurement and statistical norms.<sup>33</sup> In this body of research, the interpretive dimensions of data collection and maintenance can, and should, become a source of study. These contemporary data practices also set out to expose ‘the double role of statistics’ both in modeling as well as shaping the world.<sup>34</sup> In the process, activists reveal how data produce reality as much as they build consensus arounds statistical claims.

Both sets of literature reveal important insights about activism through data archiving. Through a post-positivist lens, we can highlight the political importance of data as a tool to make claims about the world *and* of archives to maintain these statistical representations for our future understandings of contentious issues.

## ***2. Both Forms of Activism Address Issues of Representation***

The social and political power of data and archives is in large part due to their capacity to make people and phenomena visible. Community archives, for example, give marginalised populations an active role in creating cultural memory, either through oral histories or by collecting and stewarding documents and artifacts from community members. Such is the case of

The Miracle, a California-based queer archive and book mobile project, which formed in part to counteract the loss of LGBTQ community spaces such as book stores and nightclubs, often through gentrification.<sup>35</sup> Activist archives can also make these voices heard by asking them to take part in the records' custody, for instance by contributing metadata and context; the participatory South Asian American Digital Archive (SAADA), for example, allows contributors to upload and contextualise records from personal collections.<sup>36</sup>

Similarly, data activism gives citizens a means to take control over their own representations or those of contentious issues. Grassroots statistical accounts can, for one, make an issue visible that was formerly unaccounted for in any official capacity; an example of this is the all-volunteer Los Angeles Country Bike Coalition's annual Bike and Pedestrian count, which makes cyclists and pedestrians statistically visible for Los Angeles City and County planners making important infrastructure and transportation planning decisions.<sup>37</sup> Data activism can also augment existing official statistical accounts by working with the state to bring numbers and metrics more in line with activists' understanding of it. The Gowanus Canal Conservancy, for instance, is a community organization that contributes data to official state maps of water inflows entering the polluted canal.<sup>38</sup> Data activists can, alternately, offer a completely alternative statistical account that in does not try to reconcile with official numbers. An example of such 'agonistic data' practitioners includes the Superstorm Research Group, a part of the grassroots Occupy Sandy collective.<sup>39</sup> While a poll by the City of New York's Deputy Mayors asked about the success of evacuation orders, how long people were away from their houses, and other metrics that matter for FEMA and government agencies,<sup>40</sup> the Research Group, in contrast, surveyed residents' socioeconomic situation and preexisting crises that had been exacerbated by the weather event. The information collected by both the City and the Superstorm Research

Group resulted in two very different types of information pertaining to the same phenomenon, with the Research Group's highlighting how systemic class issues are exacerbated by extreme weather.

Activists can archive data created and collected by themselves, in a fashion similar to community archives. Archival data activism can also follow another approach: it can re-contextualise already existing official data by decentralizing the archive. Such work relates closely to electronic participatory archives, which open their collections to members from disparate contexts to work with the same set of records, allowing participants to link electronic records to related public records or to the records owned by other communities, contribute and annotate records in a way that represents multiple provenance and contested views, demonstrate shared rights, and allow differential access to culturally sensitive records in a manner that respects local customs.<sup>41</sup> DataRescues engage in a similar form of archival re-appropriation: activists did not collect the data themselves but re-contextualised the data by downloading it from official websites, mirroring it on alternate servers, and adding their own metadata to the datasets through custom-built web software, as we describe in a later section. In this way the data is re-presented in a politicised fashion, as activists concerned about climate change research steward the data outside of official contexts.

### ***3. Both Forms of Activism Can Be Trustworthy and Make Legitimate Claims***

Grassroots archives provide an infrastructure of legitimacy for records that might otherwise remain in ad hoc relation to each other. Operating outside of well-known institutions, activist and community archives can still offer legitimate sources of information and memory, whether by maintaining information on a record's provenance or providing a dedicated repository. In this way, activist archives can introduce new sources of evidence into legal or political realms. Such

a role is particularly important for the documentation of human rights or labor abuses. The Documentation Center of Cambodia, for example, formed to archive documentary evidence against surviving Khmer Rouge leaders in order to make the case for a political tribunal that could redress past injustices.<sup>42</sup>

Data activists must also address issues of legitimacy within the public sphere. Data are often the result of ‘centres of calculation,’ the term Bruno Latour uses for powerful institutions that collect, store, and process information.<sup>43</sup> Government administrations, think tanks, and universities, for instance, have the financial and political resources to produce and publish expensive statistical knowledge. In turn, data activists must also deploy methodologies that are the public and other institutions consider legitimate. In order to generate claims that appeal to centers of calculation - or to challenge their authority – activists must ensure that their statistics are accepted by experts and general society.<sup>44</sup>

Civic technoscience projects serve as a model here: they deliberately involve non-experts and community-based groups in scientific questioning and data production, using devices to gather data that are robust and scientifically vetted.<sup>45</sup> ‘Bucket brigades,’ as one example, equip citizens with EPA-approved buckets to collect air samples to measure for toxic emissions. The brigades formed out of a practical necessity to address the lack of scientific tools that would allow non-scientists – particularly citizens affected by pollution – to participate in air monitoring.<sup>46</sup> In another example, activists from the Sierra Club ordered formaldehyde test kits to test trailer homes providers to those who had lost their homes in the aftermath of Hurricane Katrina.<sup>47</sup> The tests, which found unsafe formaldehyde levels, provided evidence that later awarded the mobile home owners with \$42.6 million in a 2012 class action law-suit.<sup>48</sup> Statistical claims in this way expand the voices of those who contribute to authoritative metrics on a



phenomenon; they redress the imbalance of power between government or scientific claims and those produced by grassroots efforts.

To establish legitimacy, data archivists must document that their data is untampered with and has not degraded in any way; they must use professional standards for digital archiving, including metadata that documents chain of custody and provenance and preserves context and sometimes multiple copies to prevent loss. Data activists may need to create their own software to fulfill sound archival practices. For example, EDGI volunteers, as we describe in the case study, designed custom software for data ingest and metadata in order to ensure the integrity of the archive.

#### ***4. Both Forms of Activism Engage in Community Formation***

Activist and community archives address the important question of positionality: the subjects of records, in many of these projects, become ‘participatory agents,’ thereby gaining a host of responsibilities and legal rights to those records.<sup>49</sup> In the process, a dialectic can occur: as a community stewards an archive, the archival work itself also *produces* a community of activists dedicated to negotiating questions such as inclusion, accessibility, identity, and voice. As such, the boundaries and identity of the community are always at play in a dynamic process.<sup>50</sup> Many community archives also have outreach and educational missions that do more than just assert visibility; they ultimately aim to transform society and are integral for wider movement building.<sup>51</sup>

Similar to the feminist and de-colonial goals of participatory archiving, data activism can allow the subjects of statistical records a voice in the metrics that represent them; as a result, data become less abstract and more rooted within the realm of communal practice. A community may gather around the statistical work of collecting, interpreting, visualizing, and stewarding data.

The Port Arthur, Texas, Community in-Power and Development Association (CIDA), for example, formed with the mission to track and monitor pollutants in fence-line communities nearby large oil refineries, including the Motivo Oil Plant, the largest on the North American continent. CIDA uses the data they collect to hold government and industry accountable.<sup>52</sup>

The fact that people are affected by data suggests, as Dalton and Thatcher argue, the necessity for critical attention to the positionality of those involved.<sup>53</sup> The hackathon model, for instance, is commonly used in civic data projects, but these events primarily attract highly-educated and skilled participants.<sup>54</sup> Data activism can benefit from participatory archival approaches through which communities define and develop their own categories for statistical data—seeking to ‘measure what matters’ to them—and learn about data curation and management in the process.<sup>55</sup> By collaboratively raising questions about what counts as data, what data is collected, to what ends the data is preserved and wielded, and the politics of data collection—including whose knowledge and expertise is valued - data activism can be more sensitive to cultural questions of data literacy that often privilege certain skills over others.

One example of data activism that addresses positionality and community formation is a hackathon hosted by authors in early 2015. In 2014, the deaths of Michael Brown, Eric Garner and others ignited widespread concerns about racial profiling and unnecessary use of force by police. These incidents also drew public attention to the fact that at the time, no one knew the number of people killed by police across the United States each year.<sup>56</sup> Concerned and motivated by these events, our research team downloaded and organized data on police homicides in Los Angeles County from federal and local sources, including the Justice Department, the LA Times, and the Youth Justice Coalition, a Los Angeles-based non-profit. We then opened these datasets to the public to make sense of it, inviting over fifty community members—from grassroots

organisers and activists, to interested professors and students from institutions across Los Angeles, to private citizens—to collaboratively investigate the data. Participants reflected on overarching issues of incomplete and uncorroborated data collection, as well as the processes of categorizing types of homicides. We developed projects to investigate and augment the data, by templating surveys to collect qualitative accounts or creating visualisations that reveal discrepancies in the extant datasets. This intervention is an example of how interrogate data within diverse communities gathered around shared matters of concern.

### ***5. Both Forms of Activism Can Persist Independently of Traditional Institutional Contexts***

Autonomy, according to many archival theorists, is essential to the success of activist and community archives.<sup>57</sup> Capitalist democratic institutions, in their inability to address society's inequalities, create the need for archival activism. Caswell, et al. state, 'Memory work should be dangerous';<sup>58</sup> it should disturb, not reinforce, hierarchies of power. So long as the past and present marginalization of and violence towards particular minority communities remain central to mainstream institutions, whether corporations, universities or government-funded historical archives, such institutions cannot be counted on to memorialise meaningfully on behalf of these voices. By remaining independent from formal institutions, activist archives create 'safe spaces' that make a statement about how entrenched institutions play a role in their political necessity in the first place.<sup>59</sup>

Likewise, in the realm of data activism, activists may collect data that is politically unpalatable to governments or corporate institutions. An example of the stewardship of such contested data are online data portals of police homicides in the United States, such as Fatal Encounters and killedbypolice.net, which exist because the federal government routinely fails to collect data on these incidents.<sup>60</sup> The Fatal Encounters project, for instance, operates outside of

institutional purview as a volunteer-run project; it was paid for by its founder, journalist Brian Burghardt, from its inception until 2014, when it started to receive grant and crowdfunding. While Burghardt is funded by grants to oversee the project, the brunt of the work still relies on volunteers who submit content through the organisation's website. The work is nevertheless robust, and it has become a go-to source on police officer-involved homicides for the news media in the U.S.<sup>61</sup> These websites demonstrate the crucial role of independent civil society to give legibility to an issue and hold government violence to account.

### **Case Study: The Environmental Data and Governance Initiative (EDGI)**

Data activism and archival activism both recognize the power of information - whether archival records or statistical data - to make issues visible and people's claims legitimate. We believe that both forms of activity have much to learn from each other and that the need to steward data over the long term will only grow, particularly, as the tools to produce statistical data become more widespread and ad hoc.

In this section, we examine how these two forms of activism can overlap in practice, by focusing on a particular network of activists, The Environmental Data & Governance Initiative (EDGI). EDGI combines archival and data activism to devise new software and protocols for database archiving; the initiative is a collaborative network that investigates potential threats to the scientific research infrastructure integral to strong environmental and energy policy in the United States and around the world. For this research, we spoke to three members closely involved in EDGI's archival software and workflow development in order to gather details of the processes involved; both authors have also been involved in EDGI projects, including hosting a DataRescue and taking part in EDGI's website monitoring and data justice projects.<sup>62</sup>

EDGI began in November 2016 in response to the election of Donald Trump to the U.S. presidency. EDGI's founders included Canadians who had recent memories of former Prime Minister Stephen Harper's climate change skepticism and the massive reduction in public funding of Canada's environmental research and outreach. Harper had ordered the physical destruction of materials from scientific libraries around the country and silenced government climate scientists. Internationally, scholars shared similar concerns that Trump's ideological position on climate change would also result in the removal of scientific information from public access.

EDGI formed with several goals in mind, including the design of web-based tools for data archiving and website monitoring, fostering an international network of researchers, and hosting public events in an effort to keep public environmental data accessible.<sup>63</sup> EDGI's archival work focuses on inaccessibility; while scientific data is not being destroyed (as per law), the concern is that rather that information on climate change is becoming less accessible under the current administration. For instance, EDGI has released several reports describing how information on climate change has disappeared across federal websites and can now only be found in the Obama-era archives. While no datasets have been removed, EDGI noted 'shifts in climate policies, how climate change is described, how adaptation and mitigation efforts are framed, and overall public access to information on climate change.'<sup>64</sup>

In order to preempt any attempt by the administration to undermine climate change research, EDGI's archiving work first involved the coordination of 'DataRescue' events around the United States, organised in collaboration with the Internet Archive and two other data archiving projects, the University of Pennsylvania's DataRefuge and Climate Mirror. DataRescues invited volunteer participants together to copy federal scientific datasets,

documents, and webpages; over 30 of these occurred from December 2016 until June 2017. In tandem with the archiving efforts below, DataRescues also included workshops on the political nature of data and possible impacts of the DataRescue interventions.

Most activity at DataRescues entailed webpage archiving, notably of content authored by agencies such as the Environmental Protection Agency. For this work EDGI teamed with the Internet Archive's End of Term (EoT) project, which routinely archives .gov websites during periods of a presidential transitions. EDGI volunteers focused on ensuring that EoT's web crawlers comprehensively collected the pages of agencies devoted to energy and environmental research, such as NOAA and the EPA. These basic crawlers, however, encountered many intractable webpages, web assets, interfaces and datasets that they could not crawl.

EDGI participants were then left with a set of *uncrawlable* data in need of archiving. To this end, volunteers designed an open source web application called Archivers.space, a project management tool that uses archival principles to manage the dataset's full lifecycle.

Archivers.space organises a dataset from its initial uploading to a server through multiple stages of research and vetting by participants. In one early step, participants run a python script that generates checksums – an activity that EDGI participants call 'bagging'. In another step, participants create a .zip file of both the data and any descriptive metadata, such as its broader web context. In a final step, archivists fully describe the dataset's chain of custody, context, and provenance (a step that EDGI plans to automate in future iterations). Once these steps are complete, the dataset is published on the public website DataRefuge.org (maintained by DataRefuge), where the general public can use it.

**[Figure 1. Archivers.space. A list of datasets that have been successfully harvested, bagged, and described.]**

**[Figure 2. Descriptive metadata of a single dataset within Archivists.space.]**

**[Figure 3. DataRefuge homepage]**

By decentralizing federal datasets, EDGI deploys a sound records management principle: records are less vulnerable when copies exist in multiple places. Because the data do not entail privacy concerns, EDGI uses open licenses to encourage mirroring so that a dataset exists on multiple servers, rather than attached to one central institution. With careful archival metadata the copies can remain updated, citable and have a chain of custody. EDGI's end goal and methods are, in this way, similar to those of the decentralized library repository project LOCKSS ('Lots of Copies Keeps Stuff Safe') started at Stanford University, a project that focuses on decentralizing electronic copies of library holdings so that if any one copy is vulnerable to loss, other copies exist on other servers.<sup>65</sup>

**[Figure 4. EDGI's data archiving workflow (image created by the authors)]**

Archiving data does not necessarily get it in the hands of the researchers, policymakers, and climate activists who use it in their daily work; digital data requires a well-designed interface to make it findable and usable by broader audiences. For this reason, since the summer of 2017, EDGI has moved on from collecting data to investigating how decentralized web infrastructure can make data archiving the collective project of disparate stakeholders interested in environmental and climate change, including scientists, journalists, and activists.<sup>66</sup> Data Together is a model for distributed, peer-to-peer data stewardship that uses the open source interplanetary file system (IPFS) protocol, developed at Protocol Labs, to replace the current HTTP protocol that underlies the World Wide Web.<sup>67</sup> Unlike HTTP, which stores and retrieves data by its *location* at a particular webpage, IPFS uses a hash function that uniquely identifies a piece of *content*<sup>68</sup> - the difference between using an address as your unique identifier, rather than

a national ID number.<sup>69</sup> By linking to content rather than location, IPFS fixes the problem of web links that break if content moves to another site.<sup>70</sup> EDGI's collaboration with Data Together builds decentralized web infrastructure and tools 'that allow communities to work cooperatively, share responsibilities, and reinforce each other so that data is accessible to all, immediately discoverable, easily verifiable, and robustly preserved.'<sup>71</sup>

Decentralization, however, imparts its own set of archival problems. In a centralized model, data is authorized while the central repository checks and stores the data. In the decentralized web, that dual problems of verification and authority are trickier, as are the attendant problems of locating data and storing it reliably.<sup>72</sup> IPFS, to this end, provides *verification* - it ensures that the collectively stewarded data is correct and untampered with - and high performance, which is useful if one is working with and analyzing huge datasets. IPFS provides these benefits by issuing each file and all of the blocks within it a unique fingerprint called a cryptographic hash. Using this hash to check content, IPFS removes duplicate copies across the network and tracks version history for every file. Each network node stores only content it is interested in, along with metadata that indicates provenance and ownership. When a user looks up files, she asks the network to find nodes storing the content signed by a unique hash.<sup>73</sup> Data Together's decentralized approach hopes to lay a foundation for a reliable open data commons.

To conclude, EDGI's work exemplifies the characteristics of both activist and data archives, as outlined above. EDGI shows the political importance of supposedly neutral scientific data along with the power of archival work to safeguard politically vulnerable information. EDGI's work also contends with questions of representation and voice: its archiving practices involve volunteers with no professional background in science nor the archival profession; their



work has nonetheless played a significant role in exposing efforts to suppress climate change information at the federal level (in fact, several of EDGI's reports have contributed to articles in major news outlets).<sup>74</sup> The Initiative's work is also dedicated to the integrity of the data sets over the long term by using metadata practices that ensure quality and provenance. As a result of this archival practice, a sustained, healthy community of volunteers and activists now hold regular meetings over the Web, using Slack and video conferencing, and gather in-person at academic conferences and community events around the world. Finally, EDGI has become a sustainable non-profit that operates independently of larger institutions, such as Universities or government agencies; its data archive is, therefore, also a resource that is necessarily independent of whims of politically motivated climate denial. Furthermore, its custom-built software allows EDGI to avoid commercial platforms that could undermine the archiving process through conflicting claims of ownership.

## **Conclusion**

Thinking of data archiving as a political act alters the conditions of possibility for agency in an increasingly data-driven society. While data holds immense power when applied from the top-down, the examples above show that there are ways this power can be wielded from the grassroots level. These examples also show the need for researchers to consider how two different streams of theoretical literature – one on archival activism, another on data activism – can productively learn from each other and influence communities of practice. Archival activism, for one, can consider the specific concerns of archiving statistical representations on matters of justice, cultural authority, and human rights. Data activism, in turn, can draw from archival principles to consider the technological and administrative needs of stewarding their data over

the long term – opening new avenues to pursue in mobilizing power to fight for a more just future.

EDGI's case study shows that there are clear affinities between activist archival practices and data activism, as both focus on including communities of interested stakeholders in building representations that assert their points of view. EDGI's DataRescues focused on data archiving but also encouraged thinking through political questions of *what* gets represented, *how* and for *whom*. Moreover, the case demonstrates how data activism can draw from archival activism literature to think more precisely about problems of authenticity, stewardship, autonomy, and trust over the long term in a networked environment. Answering these questions can inform how data activists frame the political necessity for the work they engage in, the infrastructures they choose to use to manage it, and the ways they partner with stakeholders and institutions.

### **Acknowledgements**

Special thanks to Joan Donovan, volunteers for the Environmental Data Governance Initiative, and Stanford's Digital Civil Society Lab.

### **Endnotes**

---

<sup>1</sup> Morgan Currie, Joan Donovan, and Britt S. Paris, 'Preserving for a More Just Future: Tactics of Activist Data Archiving,' in Usha Mujoo Munshi and Neeta Verma (ed.), *Data Science Landscape: Towards Research Standards and Protocols*, Studies in Big Data 38, Springer, Singapore: 2018, pp. 67–78.

<sup>2</sup> A Gilliland and S McKemmish, 'Recordkeeping Metadata, the Archival Multiverse, and Societal Grand Challenges,' in *Proceedings of the 2012 International Conference on Dublin Core and Metadata Applications*, DCM12. pp. 106–115, <http://dl.acm.org/citation.cfm?id=2771179.2771192>.

<sup>3</sup> M Caswell, 'Toward a Survivor-Centered Approach to Records Documenting Human Rights Abuse: Lessons from Community Archives,' *Archival Science* 14, no. 3–4, October 1, 2014, pp. 307–22, <https://doi.org/10.1007/s10502-014-9220-6>.

<sup>4</sup> Lisa Gitelman, (ed.), *Raw Data Is an Oxymoron*, MIT Press, Cambridge, MA, 2013.; M Buckland, 'Information As Thing,' *JASIST* 42, no. 5, 1991, pp 351–60.

<sup>5</sup> We ultimately see the distinction between records and data as one supplied by the two disciplines that study these, and by how practitioners themselves generally describe their work –

---

for instance, archival activists talk about the objects of their work as records, not data, while data activists talk about their efforts in terms of data, not records.

<sup>6</sup> Gilliland and McKemmish, 'Recordkeeping Metadata, the Archival Multiverse, and Societal Grand Challenges.' In this article, we assume that the archivist's role includes both 'end of life' activities as well as active engagement with the design of recordkeeping strategies, such as database design.

<sup>7</sup> A Flinn and B Alexander, 'Humanizing an Inevitability Political Craft: Introduction to the Special Issue on Archiving Activism and Activist Archiving,' *Archival Science* 15, no. 4, December, 2015, pp. 329–35, <https://doi.org/10.1007/s10502-015-9260-6>.

<sup>8</sup> Ibid.; Luciana Duranti and Patricia C Franks, 'Community Archives,' in *Encyclopedia of Archival Science*, Rowman & Littlefield, 2015, pp. 145–49; M Caswell, M Cifor, and MH Ramirez, 'To Suddenly Discover Yourself Existing: Uncovering the Impact of Community Archives,' *The American Archivist* 79, no. 1, June, 2016, pp. 56–81, <https://doi.org/10.17723/0360-9081.79.1.56>.

<sup>9</sup> A Sellie et al., 'Interference Archive: A Free Space for Social Movement Culture,' *Archival Science* 15, no. 4, December, 2015, pp. 454, <https://doi.org/10.1007/s10502-015-9245-5>.

<sup>10</sup> Today the Mazer is held by the University of California, Los Angeles, in a case of an archive shifting from the hands of an autonomous, grassroots community to a formal institution. Please see <http://www.mazerlesbianarchives.org>.

<sup>11</sup> Sam Walker, 'Black Cultural Archives,' in Alison Donnell (ed.), *Companion to Contemporary Black British Culture*, Routledge, London, 2002, pp. 45.

<sup>12</sup> Christine L Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, MA, 2015.

<sup>13</sup> Gitelman, *Raw Data Is an Oxymoron*.

<sup>14</sup> C Dalton and J Thatcher, 'What Does a Critical Data Studies Look like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data,' *Society & Space*, 2014, <http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/>.

<sup>15</sup> I Bruno, E Didier, and T Vitale, 'Statactivism: Forms of Action between Disclosure and Affirmation,' *The Open Journal of Sociopolitical Studies, Partecipazione E Conflitto*, 7, no. 2, 2014, pp. 198–220.

<sup>16</sup> S Milan, 'Data Activism as the New Frontier of Media Activism,' SSRN Scholarly Paper, Social Science Research Network, January 31, 2016, <https://papers.ssrn.com/abstract=2882030>.

<sup>17</sup> Y Kawano, et al., 'A Map for the Future: Measuring Radiation Levels in Fukushima, Japan,' in 2012 IEEE Global Humanitarian Technology Conference, 2012, pp. 53–58, <https://doi.org/10.1109/GHTC.2012.18>.

<sup>18</sup> M Currie et al., 'The Conundrum of Police Officer-Involved Homicides: Counter-Data in Los Angeles County,' *Big Data & Society* 3, no. 2, December, 2016, <https://doi.org/10.1177/2053951716663566>.

<sup>19</sup> P O'Brien, 'Brown Moses: The British Blogger Tracking Syrian Arms,' Channel 4 News, March 30, 2013, <https://www.channel4.com/news/brown-moses-blog-syria-arms-weapons-croatia>.

<sup>20</sup> L Dencik, A Hintz, and J Cable, 'Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism,' *Big Data & Society* 3, no. 2, December 1, 2016, <https://doi.org/10.1177/2053951716679678>.

- 
- <sup>21</sup> Ian Hacking, 'Prussian Numbers 1860-1882,' in *The Probabilistic Revolution*, Volume MIT Press, Cambridge, MA, 1987, pp. 377–94; Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, Harvard University Press, 2002.; I Hacking, 'Making Up People,' *London Review of Books*, 2006; I Bruno, E Didier, and T Vitale, 'Statactivism: Forms of Action between Disclosure and Affirmation.'
- <sup>22</sup> Dalton and Thatcher, 'What Does a Critical Data Studies Look like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data.'
- <sup>23</sup> Bruno, Didier, and Vitale, 'Statactivism: Forms of Action between Disclosure and Affirmation,' p. 199.
- <sup>24</sup> Steven Epstein, *Impure Science: AIDS, Activism, and the Politics of Knowledge*, University of California Press, Berkeley, CA, 1996.
- <sup>25</sup> C Lanius, 'Fact Check: Your Demand for Statistical Proof Is Racist,' *Cyborgology*, January 12, 2015, <https://thesocietypages.org/cyborgology/2015/01/12/fact-check-your-demand-for-statistical-proof-is-racist/>.
- <sup>26</sup> S Marche, 'Literature Is Not Data: Against Digital Humanities,' *Los Angeles Review of Books*, October 28, 2012, <https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/>.
- <sup>27</sup> R Holifield, 'Neoliberalism and Environmental Justice in the United States Environmental Protection Agency: Translating Policy into Managerial Practice in Hazardous Waste Remediation,' *Geoforum*, 35, no. 3, May 1, 2004, pp. 285–97, <https://doi.org/10.1016/j.geoforum.2003.11.003>.
- <sup>28</sup> Bruno, Didier, and Vitale, 'Statactivism: Forms of Action between Disclosure and Affirmation'; Desrosières, *The Politics of Large Numbers*.
- <sup>29</sup> H Zinn, 'Secrecy, Archives, and the Public Interest,' *Midwest Archives* 2, no. 2, 1977, pp. 14–26.
- <sup>30</sup> Henry Jenkinson, *Selected writings of Sir Hilary Jenkinson*. Alan Sutton Pub, Gloucester England, 1980.
- <sup>31</sup> T Cook, 'Archival Science and Postmodernism: New Formulations for Old Concepts,' *Archival Science* 1, no. 1, 2001, pp. 3–24.
- <sup>32</sup> CB Trace, 'What Is Recorded Is Never Simply 'What Happened: Record Keeping in Modern Organizational Culture,' *Archival Science* 2, no. 1–2, March 1, 2002, pp. 137–59, <https://doi.org/10.1007/BF02435634>.
- <sup>33</sup> Hacking, 'Prussian Numbers 1860-1882'; P Agre, 'Living Data,' *Wired*, November 1996, <http://archive.wired.com/wired/archive/2.11/agre.if.html>; Geoffrey C Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, MIT Press, 2000; Hacking, 'Making Up People'; R Kitchin and TP Lauriault, 'Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work,' *Social Science Research Network*, July 30, 2014, <https://papers.ssrn.com/abstract=2474112>.
- <sup>34</sup> Desrosières, *The Politics of Large Numbers*.
- <sup>35</sup> R Crooks, I Contreras, and K Besser, 'Herstory Belongs to Everybody or The Miracle: A Queer Mobile Memory Project,' *InterActions: UCLA Journal of Education and Information Studies* 11, no. 2, January 1, 2015, <https://escholarship.org/uc/item/4d67f235>.
- <sup>36</sup> M Caswell and S Mallick, 'Collecting the Easily Missed Stories: Digital Participatory Microhistory and the South Asian American Digital Archive,' *Archives and Manuscripts* 42, no. 1, January 2, 2014, pp. 73–86, <https://doi.org/10.1080/01576895.2014.880931>.
- <sup>37</sup> See <https://la-bike.org/our-work/bike-ped-count/>.

- 
- <sup>38</sup> SA Wylie et al., 'Institutions for Civic Technoscience: How Critical Making Is Transforming Environmental Research,' *The Information Society* 30, no. 2, March 1, 2014, pp. 116–26, <https://doi.org/10.1080/01972243.2014.875783>.
- <sup>39</sup> Morgan E Currie, 'The Data-Fication of Openness - The Practices and Policies of Open Government Data in Los Angeles', UCLA, 2016, <https://escholarship.org/uc/item/4xn3832z>.
- <sup>40</sup> Max Liboiron, 'Disaster Data, Data Activism: Grassroots Responses to Representing of Superstorm Sandy,' in Julia Leyda and Diane Negra (ed.), *Extreme Weather and Global Media*, Routledge, London, 2015, pp. 144–62.
- <sup>41</sup> I Huvila, 'Participatory Archive: Towards Decentralised Curation, Radical User Orientation, and Broader Contextualisation of Records Management,' *Archival Science* 8, no. 1, March 1, 2008, pp. 15–36, <https://doi.org/10.1007/s10502-008-9071-0>.
- <sup>42</sup> Michelle Caswell, *Archiving the Unspeakable: Silence, Memory, and the Photographic Record in Cambodia*, University of Wisconsin Press, Madison, WI, 2014.
- <sup>43</sup> Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society* MIT Press, Cambridge, MA, 1987, p. 215.
- <sup>44</sup> P Baudot, 'Who's Counting? Institutional Autonomy and the Production of Activity Data for Disability Policy in France (2006-2014),' *Partecipazione E Conflitto* 7, no. 2, July 7, 2014, pp. 294–313, <https://doi.org/10.1285/i20356609v7i2p294>.
- <sup>45</sup> Wylie et al., 'Institutions for Civic Technoscience.'
- <sup>46</sup> G Ottinger, 'Buckets of Resistance: Standards and the Effectiveness of Citizen Science,' *Science, Technology, & Human Values* 35, no. 2, March 1, 2010, pp. 244–70, <https://doi.org/10.1177/0162243909337121>.
- <sup>47</sup> H Smith, 'People Are Still Living in FEMA's Toxic Katrina Trailers — and They Likely Have No Idea', *Grist*, August 27, 2015, <https://grist.org/politics/people-are-still-living-in-femas-toxic-katrina-trailers-and-they-likely-have-no-idea/>.
- <sup>48</sup> The Associated Press, 'FEMA Mobile Home Manufacturers Agree to \$2.6 Million Settlement,' accessed January 31, 2018, [http://www.nola.com/katrina/index.ssf/2011/01/fema\\_mobile\\_home\\_manufacturers.html](http://www.nola.com/katrina/index.ssf/2011/01/fema_mobile_home_manufacturers.html).
- <sup>49</sup> Gilliland and McKemmish, 'The Role of Participatory Archives in Furthering Human Rights, Reconciliation and Recovery,' pp. 82.
- <sup>50</sup> Crooks, et al., 'Herstory Belongs to Everybody or The Miracle.'
- <sup>51</sup> A Flinn, M Stevens, and E Shepherd, 'Whose Memories, Whose Archives? Independent Community Archives, Autonomy and the Mainstream,' *Archival Science* 9, no. 1–2, June 2009, pp. 71–86, <https://doi.org/10.1007/s10502-009-9105-2>.
- <sup>52</sup> Please see <http://www.cidainc.org>.
- <sup>53</sup> Dalton and Thatcher, 'What Does a Critical Data Studies Look like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data.''
- <sup>54</sup> T Lodato, 'Ethnography of Civic Participation: The Difficulty of Showing Up Even When You Are There,' *EPIC*, May 2, 2016, <https://www.epicpeople.org/ethnography-of-civic-participation/>.
- <sup>55</sup> E Moore et al., 'Measuring What Matters,' Pacific Institute, May 2009, <https://pacinst.org/publication/measuring-what-matters/>.
- <sup>56</sup> R Fischer-Baum, 'Nobody Knows How Many Americans the Police Kill Each Year,' *FiveThirtyEightPolitics*, August 19, 2014, <http://fivethirtyeight.com/features/how-many-americans-the-police-kill-each-year/>.

- 
- <sup>57</sup> A Flinn, 'Migrations, Disputed Heritages and Multicultured Identities: Archives in Post-Colonial Societies.,' *Archives and Manuscripts* 36, no. 2, 2008, pp. 54–75; M. L. Caswell, D. Johnson, and J. Drake, 'From Cape Town to Chicago to Colombo and Back Again: Towards a Liberation Theology for Memory Work,' Nelson Mandela Foundation, January 1, 2017, <https://www.nelsonmandela.org/news/entry/reflections-from-the-2016-mandela-dialogues>.
- <sup>58</sup> Caswell, Johnson, and Drake, 'From Cape Town to Chicago to Colombo and Back Again.'
- <sup>59</sup> Ibid.
- <sup>60</sup> Currie et al., 'The Conundrum of Police Officer-Involved Homicides.'
- <sup>61</sup> E Markowitz, 'Meet the Man Who Spends 10 Hours a Day Tracking Police Shootings,' GQ, July 8, 2016, <https://www.gq.com/story/fatal-encounters-police-statistics-interview>.
- <sup>62</sup> A Khan, 'Fearing Climate Change Databases May Be Threatened in Trump Era, UCLA Scientists Work to Protect Them.' *Los Angeles Times*, January 21, 2017, <http://www.latimes.com/science/sciencenow/la-sci-sn-climate-change-data-20170121-story.html>.
- <sup>63</sup> EDGI, 'Introducing the Environmental Data and Governance Initiative — Report,' EDGI, accessed January 31, 2018, <https://envirodatagov.org/publication/introducing-edgi/>.
- <sup>64</sup> T Rinberg, et al. EDGI, 'Changing the Digital Climate,' January 2018, <https://100days.envirodatagov.org/changing-digital-climate/>.
- <sup>65</sup> 'LOCKSS | Lots of Copies Keep Stuff Safe,' accessed April 12, 2018, <https://www.lockss.org/>.
- <sup>66</sup> EDGI, 'Next Phase of Archiving: Data Together,' EDGI, July 5, 2017, <https://envirodatagov.org/archiving-data-together/>.
- <sup>67</sup> 'LOCKSS | Lots of Copies Keep Stuff Safe.'
- <sup>68</sup> To find the content, IPFS uses 'deterministic querying,' developed by qir.io, designed for locating information in content-addressed network systems.
- <sup>69</sup> V Aurora, 'The Code Monkey's Guide to Cryptographic Hashes for Content-Based Addressing,' November 12, 2007, <http://valerieaurora.org/monkey.html>.
- <sup>70</sup> Protocol Labs, 'IPFS Is the Distributed Web,' IPFS, accessed April 11, 2018, <https://ipfs.io/>.
- <sup>71</sup> Data Together, Start Here! Discussion for Data Together: Community Patterns for the Decentralized Web, 2017, Data Together, 2018, <https://github.com/datatogether/datatogether>.
- <sup>72</sup> Ibid.
- <sup>73</sup> IPFS also resolves issues of *authority* by using decentralized trust mechanisms based on deterministic querying, in which the content address can be computed independently by anyone for any content simply by running the hash function; they do not require a centralized authority to coordinate the assignment of labels or addresses.
- <sup>74</sup> S Knutson et al. EDGI, 'EDGI Annual Report 2018', March 2018, <https://envirodatagov.org/publication/edgi-annual-report-2018/>. pps. 7, 23.